

## **WHAT IS CLAIMED IS:**

1. A method for identifying the candidate proteins useful as anti- infectives, which comprises:
  - i) calculating computationally the different sequence based attributes from all the protein sequences of the selected pathogenic organisms.
  - ii) clustering computationally all the proteins of a genome based on these sequence-based attributes using Principle Component Analysis.
  - iii) identifying computationally the outlier proteins sequences which are excluded from the main cluster.
  - iv) matching the outlier protein sequences with the protein sequences in various databases.
  - v) selecting the unique outlier protein sequences not homologous to any of the protein sequences searched above.
  - vi) validating computationally the protein sequences as anti-infectives by comparing with the known protein sequences that are biochemically characterized in the pathogen. genome.
2. A method claimed in claim 1 wherein, the protein sequence data is taken from any organism, specifically but not limited to organisms such as *B.burgdorfei*, *C.jejuni*, *C.pneumoniae*, *C.trachomatis*, *H.influenzae*, *H.pylori*, *L.major*, *M.genetalium*, *M.pneumoniae*, *M.tuberculosis*, *N.meningitis*, *P.aeruginosa*, *P.falciparum*, *R.prowazekii*, *T.pallidum*, *V.cholerae*.

3. A method claimed in claim 1 wherein different sequence-based attributes used for identification of candidate anti-infective proteins are selected from the group comprising of fixed protein and variable protein attributes.

4. A method claimed in claim 1 wherein the fixed protein attributes are selected from the group comprising of percentage of charged amino acids, percentage hydrophobicity, distance of protein sequence from a fixed reference frame, measure of dipeptide complexity of protein, and measure of hydrophobic distance from a fixed reference frame.

5. A method as claimed in claim 3 wherein the variable attribute is the distance of the protein sequence from a variable reference frame.

6. A method as claimed in claim 1, wherein the cluster analysis is carried out by Principle Analysis Technique using correlation coefficient between the attributes.

7. A method as claimed in claim 1, wherein the steps I to iv and vi are performed computationally.

8. A method as claimed in claim 1, wherein the clustering of the proteins is based upon analysis of sequence attributes instead of sequence pattern linked to biochemical functions.

9. A method as claimed in claim 1, wherein the unique outlier protein sequences non-homologous to the known anti-infective sequences specifically in the following pathogens

but not limited to, such as *B.burgdorfei*, *C.jejuni*, *C.pneumoniae*, *C.trachomatis*, *H.influenzae*, *H.pylori*, *L.major*, *M.genetalium*, *M.pneumoniae*, *M.tuberculosis*, *N.meningitis*, *P.aeruginosa*, *P.falciparum*, *R.prowazekii*, *T.pallidum*, *V.cholerae*.

10. A method as claimed in claim 1 , wherein the unique outlier sequences obtained by the method of invention that can serve as potential anti-infective candidates as listed in Table 1 and list 1.

11. A method as claimed in claim 1 , wherein The unique outlier hypothetical protein sequences from pathogenic genomes that can serve as anti-infective candidates listed in Table 2.

12. A method as claimed in claim 1 , wherein the genes encoding the unique proteins useful as anti-infectives.

13. A method as claimed in claim 1 , wherein the computer system comprises a central processing unit, executing DISTANCE program, clustering of the protein sequences based on different attributes using by Principle Component Analysis, all stored in a memory device accessed by CPU , a display on which the central processing unit displays the screens of the above mentioned programs in response to user inputs; and a user interface device.

14. A method as claimed in claim 1 , wherein the unique outlier hypothetical protein sequences from pathogenic genomes that can be used for diagnostic purpose.

15. A method as claimed in claim 1 , wherein the unique outlier hypothetical protein sequences from pathogenic genomes that can be used as vaccine candidates.

16. A method as claimed in claim 1 , wherein The unique outlier hypothetical protein sequences from pathogenic genomes that can be used for therapeutic purposes.

17. Unique outlier protein sequences non-homologous to the known anti-infective sequences specifically in the following pathogens but not limited to such as as *B.burgdorfei*, *C.jejuni*, *C.pneumoniae*, *C.trachomatis*, *H.influenzae*, *H.pylori*, *L.major*, *M.genetalium*, *M.pneumoniae*, *M.tuberculosis*, *N.meningitis*, *P.aeruginosa*, *P.falciparum*, *R.prowazekii*, *T.pallidum*, *V.cholerae*.

18. Unique outlier protein sequences as claimed in claim 17, wherein the sequences obtained by the method of invention that can serve as potential anti-infective candidates as listed in Table1 and List.

19. Unique outlier hypothetical protein sequences as claimed in claim 17, wherein the sequences from pathogenic genomes that can serve as anti-infective candidates listed in Table2.